Entropic Security and the Encryption of High Entropy Messages

Yevgeniy Dodis
New York University
dodis@cs.nyu.edu

Adam Smith
Massachusetts Insitute of Technology
asmith@theory.csail.mit.edu

September 1, 2004

Abstract

Russell and Wang [22] recently introduced an elegant, information-theoretic notion called *entropic security* of encryption: they required that the cipher text leak no predicate of the plaintext (similar to semantic security [10]) but only as long as the distribution on messages has high entropy from the adversary's point of view. They show that this notion of security can be achieved with very short keys for entropically rich message spaces. Canetti et al [6, 7] had previously constructed hash functions which satisfy a similar entropic security condition. The output of such hash function leaks no partial information about the input, provided the input has sufficiently high entropy.

This paper studies entropic security in general, and its application to the encryption of high-entropy messages.

- We elucidate the notion of entropic security. Our results apply to all entropically-secure primitives, including both encryption and hash functions. We strengthen the formulation of [6, 7, 22]: we require that an entropically-secure primitive leak no function whasoever of its input, while [6, 7, 22] focus only on *predicates*. This stronger formulation is much closer to the original notion of semantic security [10]. We also show that entropic security is equivalent to *indistinguishability* on pairs of input distributions of sufficiently high entropy. This equivalence generalizes the conventional equivalence between indistinguishability and semantic security [10]. Indistinguishability, in turn, is related to *randomness extraction* from non-uniform distributions [20].

  The proof of the equivalence of hiding predicates to hiding all functions is quite involved, and requires very different techniques from those of [10].

- We use the equivalence above, and the connection to randomness extraction, to prove several new results on entropically-secure encryption. We obtain:

  1. Two general frameworks for constructing entropically secure encryption schemes, one based on expander graphs and the other on XOR-universal hash functions. These schemes generalize the schemes of Russell and Wang, yielding simpler constructions and proofs as well as improved parameters. The best scheme uses a key of length $k = n - t + 2\log\left(\frac{1}{\epsilon}\right) + O(1)$, where ϵ is a measure of leakage.
  2. Lower bounds on the key length k for entropic security and indistinguishability. In particular, we show near tightness of Russell-Wang constructions: $k > n - t$. (In fact, for a large class of schemes $k \geq n - t + \log\left(\frac{1}{\epsilon}\right)$.)

1 Introduction

If X and Y are random variables, the statement "Y leaks no information about X" is normally formalized by requiring that X and Y be almost statistically independent. Equivalently, one can require that the Shannon mutual information $\mathbf{I}(X, Y)$ be very small. In this work we study situations where information leakage is unavoidable—that is, $\mathbf{I}(X, Y)$ is large—yet we still want a guarantee that no *useful* information about X is leaked by Y, even to a computationally unbounded adversary.

Consider an alternative notion of security, inspired by semantic security of encryptions [10]. We say Y *hides all functions of* X if for every function f, it is nearly as hard to predict $f(X)$ given Y as it is without Y, regardless of the adversary's computing power. If $Y = \mathcal{E}(X)$ for some probabilistic map $\mathcal{E}()$ (for example, an encryption scheme), then we say the map $\mathcal{E}$ is *entropically secure* if $\mathcal{E}(X)$ hides all functions of X whenever X has sufficiently high entropy.

A seemingly weaker variant of this definition has produced surprising results in at least two contexts so far: Canetti et al [6, 7] constructed hash functions whose outputs leak no partial information about the input. Russell and Wang [22] showed how one can construct entropically-secure symmetric encryption schemes with keys much shorter than the length of the input, thus circumventing Shannon's famous lower bound on key length.

Our contributions can be divided into two areas.

- We elucidate the notion of entropic security. (Our results apply to all entropically-secure primitives, including both encryption and hash functions.) We provide two new variants on entropic security, one closer in spirit to semantic security of encryptions [10], and the other along the lines of indistinguishability of encryptions. The proofs that these various notions are equivalent give us new tools for handling entropic security and highlight a relationship with "randomness extraction" from non-uniform distributions.
- We use the connection to randomness extraction to obtain new constructions and lower bounds for encryption of high-entropy messages with a short key. We simplify the constructions and proofs of Russell and Wang. We obtain simpler schemes with slightly shorter keys and much simpler proofs of security. We also prove the first lower bounds on key length of entropically secure schemes. The proofs are simple and show that the constructions of Russell and Wang (and ours) are nearly optimal.

1.1 Background

Although our general results apply to all entropically-secure primitives, we present entropic security (and our results) in the context of symmetric-key one-time encryption. Alice and Bob share a secret key K and Alice wants to securely send some message M to Bob over a public channel. M is assumed to come from some a-priori distribution on $\{0,1\}^n$ (e.g., uniform), and the goal is to compute a ciphertext E which: (a) allows Bob to extract M from E using K; (b) reveals "no information" about M to the adversary Eve beyond what she already knew. Below, we write $E \leftarrow \mathcal{E}(M, K)$ and $M = \mathcal{D}(E, K)$.

Perfect and Computational Security The first formalization of this problem came in a fundamental work of Shannon [23], who defined "no information" by requiring that M and E be independent as random variables: using information theoretic notation, $\mathbf{I}(M; E) = 0$, where $\mathbf{I}$ is the mutual information. He showed a lower bound on key length for his definition: encrypting messages of length n requires at least n bits of shared key (more formally, the Shannon entropy of the key must be at least that of message distribution: $\mathbf{H}_{sh}(K) \geq \mathbf{H}_{sh}(M)$). This bound is tight when the message is chosen uniformly from all strings of a fixed length n, since one can use a one-time pad. This bound was extended to the interactive setting by Maurer [18].

Goldwasser and Micali [10] relaxed the notion of perfect security to the *computational* setting: namely, any *efficient* Eve can extract only negligible "information" about M from E. They had to properly redefine the notion of "information", since mutual information or conditional probabilities do not make much sense in a computationally-bounded world. They suggested two now classical definitions. Consider the following, equivalent version of Shannon's definition: the encryption of any two messages yield the same distribution on ciphertexts, that is $\mathcal{E}(m_0) = \mathcal{E}(m_1)$. The first definition of Goldwasser and Micali, called *computational indistinguishability of encryptions*, generalizes this version of perfect security: they require that no efficient (polynomial-time adversary) can distinguish the encryptions of m_0 and m_1 with advantage more than ϵ over random guessing, where ϵ is some negligible quantity. Their second notion is called *semantic security*: for *any* distribution on messages M and any function $f()$, the adversary can predict $f(M)$ given $\mathcal{E}(M)$ with probability only negligibly better than she could without seeing $\mathcal{E}(M)$. The first definition is easier to work with, but the second definition seems to capture a stronger, more intuitive notion of security: for example, indistinguishability is the special case of semantic security when the message distribution M is restricted to uniform distributions over two points $\{m_0, m_1\}$. In fact, Goldwasser and Micali showed that the two definitions are equivalent. Thus, distributions with entropy 1 are in some sense the hardest to deal with for semantic security.

Statistical Security? A natural intermediate notion of security between perfect and computational security would be some kind of *statistical security*: Eve is computationally unbounded, as in the perfect setting, but can potentially recover some negligible amount of "information", as in the computational setting. At the first glance, it seems like there is no gain in this notion, no matter how we interpret "information". For example, following Shannon's approach we could require that $\mathbf{I}(M; E) \leq \epsilon$ instead of being 0. Unfortunately, Shannon's proof still implies that $\mathbf{H}_{sh}(K) \geq \mathbf{H}_{sh}(M) - \epsilon$. Similarly for indistinguishability: since $\mathcal{E}(m)$ should look almost the same for *any* fixed m, one can argue that $\mathbf{I}(E; M) = \mathbf{H}_{sh}(\mathcal{E}(M)) - \mathsf{Exp}_m \left[\mathbf{H}_{sh}(\mathcal{E}(m))\right]$ still has to be negligible, and so the key must again have entropy almost $\mathbf{H}_{sh}(M)$.

In his original work Shannon envisioned applications where Eve has a lot of uncertainty about the message. To get a pessimistic bound that $\mathbf{H}_{sh}(K) \geq n$, one only has to take M to be uniformly distributed in $\{0,1\}^n$. In fact, in the perfect setting, security against the uniform distribution implies security against *any* distribution on messages. On the other hand, the notions of indistinguishability and semantic security primarily deal with min-entropy 1 distributions, and the straightforward extension of Shannon's bound to the statistical versions of these notions *crucially uses this fact*. Thus, it is natural to ask if we can meaningfully define (statistical) semantic security and/or indistinguishability for high min-entropy distributions (say, uniform), similar in spirit to the original work of Shannon. And if yes,

1. How do these notions relate to Shannon's (statistical) notion, $\mathbf{I}(M; E) \leq \epsilon$? Most importantly, does the pessimistic bound on the key length still extend to these notions?
2. How do these notions relate to each other? Are semantic security and indistinguishability still equivalent when the message is guaranteed to have high entropy?

1.2 Entropic Security

Russell and Wang [22] introduced the idea of statistical security for encryption of high-entropy message spaces. They considered the first question above, though they focused on weakened version of semantic security. Their definition, *entropic security of encryptions for predicates*, is natural: for any distribution M of min-entropy[1] at least t and any predicate $g : \{0,1\}^n \to \{0,1\}$, Eve can

[1] The *min-entropy* of a random variable A is a measure of the uncertainty of its outcome. It is the negative logarithm of the probability that one can predict A ahead of time: $\mathbf{H}_\infty(A) = -\log(\max_a \Pr(A = a))$.

predict $g(M)$ using E only negligibly better than she could without E (here n is the message length and t is a parameter). Russell and Wang showed that Shannon's lower bound does *not* extend to this new notion: they presented two schemes beating Shannon's bound on key length, which we describe further below. Entropic security also arose earlier in work of Canetti [6] and Canetti, Micciancio and Reignold [7]. They constructed probabilistic hash functions whose output reveals no partial information about their input as long as it had sufficiently high entropy [6, 7].

We discuss a stronger version of the definition of [6, 7, 22], which requires that the adversary gain no significant advantage at predicting any function whatsoever of the input. One of our results is the equivalence of their notion of security to the one described here.

Definition 1.1 (Entropic Security). *The probabilistic map Y hides all functions of X with leakage ϵ if for every adversary $\mathcal{A}$, there exists some adversary $\mathcal{A}'$ such that for all functions f,*

$$\big|\Pr[\mathcal{A}(Y(X)) = f(X)] - \Pr[\mathcal{A}'() = f(X)]\big| \leq \epsilon.$$

The map $Y()$ is called (t, ϵ)-entropically secure if $Y()$ hides all functions of X, whenever the min-entropy of X is at least t.

1.2.1 Two Games for Measuring Information Leakage

In order to explain the relation between entropic security and the standard notion of security, we formulate two abstract games. Both games attempt to capture the meaning of the statement "$Y(X)$ leaks no information about X" by allowing an adversary to guess the value of a function $f(X)$ based on $Y = Y(X)$.

In this discussion, the pairs (X, Y) and (X', Y') are sampled independently according to the same joint distribution (i.e. $Y = Y(X)$ and $Y' = Y(X')$). Let $\mathcal{X}$ denote the range of X.

Game 1, on input y: The adversary receives y as input and outputs the description of a function $f_y : \mathcal{X} \to \{0,1\}^*$, and a string g. The adversary's gain is:

$$\Pr[f_y(X) = g \mid Y = y] - \Pr[f_y(X') = g]$$

Game 2, on input y: The adversary produces the description of a function $f : \mathcal{X} \to \{0,1\}^*$ before seeing y. The adversary then sees y and outputs a string g. The adversary's gain is:

$$\Pr[f(X) = g \mid Y = y] - \Pr[f(X') = g]$$

Now consider the adversary's expected gain in each of these games when the input Y is chosen at random.

The adversary's expected advantage in Game 1 is the standard measure of information leakage, and is well understood. It can be bounded by the statistical difference[2] between the joint distribution (X, Y) and the product of marginals (X', Y) (where X' is independent of Y). Equivalently, one can bound the mutual information $\mathbf{I}(X; Y)$.

In contrast, the adversary's advantage in Game 2 is less well understood. One gets some insight by thinking of it as a simplified, information-theoretic reformulation of semantic security of encryptions [10].

[2] The statistical difference, or total variation distance, between two probability distributions measures how easy it is to distinguish samples from one distribution from samples from the other.

1.3 Contributions of This Paper

This paper considers situations in which we simply cannot prevent an adversary from having a large advantage in Game 1—that is, we cannot prevent non-negligible Shannon information about X from being leaked by Y—and yet we can still satisfy a strong definition of secrecy by ensuring that no particular function of the input is leaked, i.e. no advantage is possible in Game 2. We provide a general technique for proving that Game 2 cannot be won—namely, it is necessary and sufficient to show that the map $Y(\cdot)$ is some kind of *randomness extractor*. We apply the technique to obtain new constructions and lower bounds for entropically-secure encryption.

A Strong Definition of Security The definition we propose (Definition 1.1) is stronger than previously studied formulations of entropic security [6, 7, 22], which only considered the adversary's ability to predict *predicates* instead of all possible functions of the secret input. (Recall that a predicate is a "yes"/"no" question, that is, a function that outputs a single bit.)

For example, the definition in terms of predicates does *not directly imply* that the adversary's chance of recovering the message itself remains low! The implication does in fact hold, but the proof is not entirely trivial. The idea is to choose a "Goldreich-Levin" predicate at random, that is to use $g_r(x) = r \odot x$ where r is a random n-bit string and $\odot$ is the binary inner product $r \odot x = \sum_i r_i x_i \mod 2$. We omit a detailed proof, since we prove much more general implications further below.

An Equivalence to Indistinguishability The key result behind all our constructions is the equivalence of the following statements:

- The map $Y()$ hides all functions of X, as long as $\mathbf{H}_\infty(X) \geq t$.
- For any two random variables X_1, X_2 which both have min-entropy at least $t-2$, the random variables $Y(X_1)$ and $Y(X_2)$ are statistically indistinguishable. (We then say $Y()$ is $(t-2)$-*indistinguishable*).

There are two main pieces to the result. First, we show that indistinguishability is equivalent to entropic security for predicates (the definition of [6, 22]). This is the easier of the two parts. Second, we show that if the adversary can gain advantage ϵ at predicting some function $f(X)$ given Y, then there exists a predicate g, which depends on f and the distribution of X such that the adversary gets nearly the same advantage at guessing $g(X)$ given Y. This last result may be of independent interest. It is an information-theoretic converse to the Goldreich-Levin hardcore bit construction, which states converts a good predictor for a particular predicate into a good predictor for some underlying function.

The equivalence provides a new application of *randomness extractors* to cryptography. Recall that an extractor takes as input an arbitrary, high entropy random source and a tiny random seed, and outputs uniformly random bits. The output bits are guaranteed to be almost uniformly distributed as long as the min-entropy of the input is above some threshold t. Randomness extractors satisfy *t-indistinguishability* by definition, and so we show that an extractor's output reveals very little information about its source. [3] The equivalence simplifies many existing proofs of security. It also strengthens previous results, since one obtains a guarantee that no function at all is leaked (as opposed to no predicate).

Finally, the result parallels—and was inspired by—the equivalence due to Goldwasser and Micali for the case of computationally secure encryption schemes [10]. The proof techniques are very different. In particular, standard techniques do not (seem to) suffice to show that predicting any function with significant advantage implies predicting a predicate.

[3] This use of extractors has been around implicitly in complexity theory for many years, for example in the use of hash functions for approximate counting. However, our abstraction, and cryptographic perspective, are novel.

Encryption of High-Entropy Messages As mentioned above, Russell and Wang [22] provided two constructions of entropically-secure encryption schemes which use short keys. Let ϵ denote the leakage—that is, the advantage which we allow the adversary. First, a deterministic scheme of the form $\mathcal{E}(M,K) = M \oplus p(K)$, which is secure only when M is uniformly distributed on $\{0,1\}^n$, where K has length only $k = 2\log n + 3\log\left(\frac{1}{\epsilon}\right) + O(1)$ and $p(\cdot)$ is some carefully designed function from k to n bits.[4] Thus, $p(K)$ could be viewed as a very sparse one-time pad which nevertheless hides any a-priori specified function $f(M)$. Second, for general min-entropy t, Russell and Wang gave a very different looking *randomized* scheme of the form $(\psi, \psi(M) + K) \leftarrow \mathcal{E}(M,K)$, where ψ is chosen at random from some special class of permutations[5] (and the addition is defined over some appropriate space). The analysis in [22] shows that this second scheme needs key length $n - t + 3\log\left(\frac{1}{\epsilon}\right) + O(1)$. While less than n for nontrivial settings of t, this key length again becomes $\Omega(n)$ when $n - t = \Omega(n)$. [22] left it open whether such dependence on $n - t$ is necessary.
We improve the results of Russell and Wang in several directions. We obtain:

1. A stronger definition of security than was previously known to hold, as well as the equivalence of security with indistinguishability.
2. Lower bounds on the key length k for entropic security and indistinguishability. In particular, we show near tightness of Russell-Wang constructions: $k > n - t$. (In fact, for a large class of schemes $k \geq n - t + \log\left(\frac{1}{\epsilon}\right)$.)
3. Two general frameworks for constructing entropically secure encryption schemes, one based on expander graphs and the other on XOR-universal hash functions. These schemes generalize the schemes of Russell and Wang, yielding simpler constructions and proofs as well as improved parameters.

The equivalence of entropic security and indistinguishability allows us to concentrate on a simpler definition, which immediately yields several benefits.

On one hand, we use it to show that the general construction of Russell and Wang is nearly optimal: *any* entropically secure scheme must have $k > n - t$. In fact, for a special case of *public-coin* schemes, where the ciphertext contains the randomness used for encryption,[6] we get an even stronger bound: $k \geq n - t + \log\left(\frac{1}{\epsilon}\right)$. The latter result is proven by relating the notion of indistinguishability to that of *randomness extractors* [20]: namely, any indistinguishable public-coin scheme almost immediately yields a corresponding extractor. Using the optimal lower bounds on extractors [21], we get our stronger bound as well. The schemes in [22] and this work are all public-coin.

On the other hand, the indistinguishability view allows us to give a general framework for constructing entropically secure encryption schemes. Specifically, assume we have a d-regular expander G on 2^n vertices V with the property that for any subset T of 2^t vertices, picking a random vertex v of T and taking a random neighbor w, we obtain an almost uniform distribution on V. Then, we almost immediately get an encryption scheme with key length $k = \log d$ which is indistinguishable for message spaces of min-entropy t. Looking at this from another perspective, the above encryption scheme corresponds to a randomness extractor which takes a source M of length n and min-entropy t, invests $\log d$ extra random bits K, and extracts n almost random bits E (with the additional property that the source M is recoverable from E and K). From this description, it is clear that the key length of this paradigm must be at least $n - t$ (which we show is required in any entropically secure encryption scheme). However, using optimal expanders we can (essentially) *achieve* this bound, and in several ways. First, using Ramanujan expanders [16], we get the best

[4] $p(\cdot)$ samples a random point $p(K)$ from an appropriate δ-biased spaces [19] (where [22] used $\delta = \epsilon^{3/2}$).

[5] Russell and Wang required a family of 3-wise indepepndent permutations.

[6] In particular, this includes all the deterministic schemes.

known construction with key length $k = n - t + 2\log\left(\frac{1}{\epsilon}\right)$. Second, using δ-biased spaces [19] (for appropriate $\delta = \delta(\epsilon, n, t)$ explained later), we get a general construction with slightly larger but still nearly optimal key length $k = n - t + 2\log n + 2\log\left(\frac{1}{\epsilon}\right)$. This last result generalizes (and slightly improves) to any value of t the special case of the uniform message distribution ($n - t = 0$) obtained by Russell and Wang [22]. Our approach also gives clearer insight as to why small-biased spaces are actually useful for entropic security.

While the deterministic constructions above are nearly optimal and quite efficient, we also observe that one can get simpler constructions by allowing the encryption scheme to be *probabilistic*. In our approach, this corresponds to having a *family* of "average case" expanders $\{G_i\}$ with the property that for any set T of size at least 2^t, picking a random graph G_i, a random v in T and taking a random neigbor w of v in G_i, we get that w is nearly uniform, *even given the graph index i*. By using any family of pairwise independent hash functions h_i (resp. permutations ψ_i) and a new variant of the leftover hash lemma [14], we get a probabilistic scheme of the form $\langle i,\ M \oplus h_i(K)\rangle$ (resp. $\langle i,\ \psi_i(M) \oplus K\rangle$) with a nearly optimal key length $k = n - t + 2\log\left(\frac{1}{\epsilon}\right)$. As a concrete example of this approach, we get the following simple construction: $\mathcal{E}(M, K; i) = (i, M + i \cdot K)$, where the local randomness i is a random element in $GF(2^n)$, $K \in \{0,1\}^k$ is interpreted as belonging to $GF(2^k) \subseteq GF(2^n)$, and addition and multiplication are done in $GF(2^n)$.

Once again, the above result (with permutations ψ_i) improves and simplifies the intuition behind the second scheme of Russell and Wang [22]. Indeed, the latter work had to assume that the ψ_i's come from a family of 3-wise independent permutations — which are more compicated and less efficient than 2-wise independent permutations (or functions) — and presented a significantly more involved analysis of their scheme.

1.4 A Caveat: Composing Entropically-Secure Constructions

A desirable property of definitions of security of cryptographic primitives is *composability*: once some protocol or algorithm has been proven secure, you would like to be able to use it as a building block in other protocols with your eyes closed—without having to worry about effects that violate the intuitive notion of security, but which are not covered by the original definition.

Composability is difficult to guarantee, since it is not clear how to translate it into a mathemetical property. There are various formalizations of composability, most notably "Universal Composability" [8] and several frameworks based on logic algebras for automated reasoning (see [13] and the references therein). Finding protocols that are provably secure in these general frameworks is difficult, and sometimes provably impossible. A more common approach is to prove that a particular definition remains intact under a few straightforward types of composition, say by proving that it is still secure to encrypt the same message many times over.

The main weakness of entropic security, as defined above, is that it does not ensure composability, even in this straightforward sense. If $Y()$ and $Y'()$ are independent versions of the same entropically-secure mapping, then the map which outputs the pair $Y(X), Y'(X)$ may be completely insecure, to the point of revealing X completely. In the case of encryption, this may mean that encrypting the same message twice is problematic. The reason is that given the first value $Y(X)$, the entropy of X may be very low, too low for the security guarantee of $Y'()$ to hold.

For example, suppose that $Y(x)$ consists of the pair M, Mx, where M is a random $\frac{3n}{4} \times n$ binary matrix M and $x \in \{0,1\}^n$. We will see later that $Y()$ is entropically secure whenever the entropy of X is close to n. However, the pair $Y(x), Y'(x)$ provides a set of $\frac{3n}{2}$ randomly chosen linear constraints on x. With high probability, these determine x completely, and so the pair $Y(), Y'()$ is insecure under any reasonable definition.

Given these issues, entropically-secure primitives must be used with care: one must ensure that the inputs truly have enough entropy for the security guarantee to hold. The requirement of entropy is natural in many situations (e.g. when the input is a password), but the issue of composability nonetheless raises a number of interesting open questions for future research.

The generality and intuitive appeal of entropic security, as well as the variety of contexts in which it has arisen, make it an important concept to understand. We hope that the present work provides a major step in this direction.

2 Entropic Security, Prediction and Indistinguishability

This section formulates the various notions of security we work with, and states the equivalences which are the main technical results of the section. Let $Y(x;R)$ be some randomized map. Here $x \in \{0,1\}^n$ is the input and R is a string of uniformly random bits, independent of m. In the case of encryption, x is a message, $Y = \mathcal{E}$ is the encryption function, and $R = \langle \kappa, i \rangle$ consists of the key and any extra randomness i used by the encryption. In the hashing setting of [6, 7], Y is the hash function and R, the randomness used by the hash. When the random string R is implicit, we will simply write $Y(x)$ instead of $Y(x;R)$.

Recall that a predicate is a "yes"/"no" question, that is, a function that outputs a single bit. Entropic security was first formulated in terms of predicates [6, 7, 22]. That definition is exactly the same as Definition 1.1, but with the class of functions restricted to have a single bit of output.

The definition may not seem quite satisfying from several points of view. First, it states only that no predicate of the input is leaked, and provides no explicit guarantees about other functions. In contrast, the original semantic security definition of Goldwasser and Micali held for all functions, not only predicates. Second, there is no guarantee that the new adversary $\mathcal{A}'()$ is polynomial time, even in the case where, say, $\mathcal{A}$ runs polynomial time and M is samplable in polynomial time. Finally, the definition is somewhat hard to work with.

We introduce two new definitions which we prove are equivalent to entropic security for predicates. First, we extend the definition of [6, 22] to hold for *all functions*, not only predicates. This is the definition discussed in the introduction (Definition 1.1).

Second, we show that the entropic security of $Y()$ is equivalent to the indistinguishability of $Y()$'s outputs on certain pairs of distributions on the inputs. This notion is inspired by that of Goldwasser and Micali [10], which required that the output of an encryption scheme on any pair of inputs (as opposed to pair of distributions over inputs) be indistinguishable by polynomial-time adversaries. One nice consequence of this equivalence is that in Definition 1.1 we can take $\mathcal{A}'() = \mathcal{A}(Y(U_n))$, where U_n is the uniform distribution on $\{0,1\}^n$. This definition is also much easier to work with, as we will see in later sections.

Definition 2.1. *A randomized map $Y()$ is* (t,ϵ)-indistinguishable *if there is a random variable G such that for every distribution on messages M over $\{0,1\}^n$ with min-entropy at least t, we have*

$$\mathbf{SD}\left(Y(M),\ G\right) \leq \epsilon.$$

Indistinguishability is stated in terms of the statistical difference $\mathbf{SD}(A,B)$ between a pair of random variables A, B. This is half the L_1 distance between the distributions, $\mathbf{SD}(A,B) \overset{def}{=} \frac{1}{2}\sum_x |\Pr[A=x] - \Pr[B=x]|$. It also has an operational meaning: given a sample from either A or B (at random), the optimal adversary's chance of correctly guessing which distribution the sample came from is exactly $\frac{1}{2} + \frac{1}{2}\mathbf{SD}(A,B)$.

We now state the main result of the section.

Theorem 2.1. *Let Y be a randomized map with inputs of length n. Then*

1. (t,ϵ)*-entropic security for predicates implies* $(t-1, 4\epsilon)$*-indistinguishability.*

2. *$(t-2,\epsilon)$-indistinguishability implies $(t,\epsilon/8)$-entropic security for* ***all functions*** *when $t \geq 2\log\left(\frac{1}{\epsilon}\right)+1$.*

Entropic security with respect to predicates is trivially implied by entropic security for all functions, and so Theorem 2.1 states that all three notions of security discussed above are equivalent up to small changes in the parameters.

Randomness Extraction and Entropic Security Taking the distribution G in Definition 2.1 to be the uniform distribution, then we recover the definition of randomness extraction—for any input distribution of high-enough entropy, the output is very close to uniform. Thus, Theorem 2.1 implies that an extractor for t-sources hides all partial information about sources of min-entropy at least $t+2$.

2.1 Proving Theorem 2.1

The remainder of this section gives an overview of the proof of Theorem 2.1.

First, some notation. Fix a distribution X on $\{0,1\}^n$. For a function $f : \{0,1\}^n \to \{0,1\}^*$, let $\mathsf{pred}_{f,X}$ be the maximum probability of any particular outcome, that is the maximum probability of predicting $f(X)$ without having any information about X:

$$\mathsf{pred}_{f,X} \overset{def}{=} \max_z \Pr[f(X)=z]$$

(When X is clear from the context, we may simply write pred_f.) We may rephrase entropic security as follows: for every function f and adversary $\mathcal{A}$, the probability of predicting $f(X)$ given $Y(X)$ is at most $\mathsf{pred}_f + \epsilon$:

$$\Pr[\mathcal{A}(Y(X)) = f(X)] \leq \mathsf{pred}_{f,X} + \epsilon$$

From Entropic Security to Indistinguishability The first statement of Theorem 2.1 is the easier of the two to prove, and we give the intuition here: given two distributions X_0, X_1, we can define a predicate $g(x)$ which captures the question "is x more likely to have come from X_0 or X_1?" If X is a equal mixture of X_0 and X_1, then the adversary which makes the maximum likelihood guess at $g(X)$ given $Y(X)$ will have success probability $\frac{1}{2}+\frac{1}{2}\mathbf{SD}\left(Y(X_0),Y(X_1)\right)$. On the other hand, with no access to $Y(X)$, the adversary can succeed with probability at most $\mathsf{pred}_P = \frac{1}{2}$. Entropic security implies that the advantage over random guessing, and hence the statistical distance, must be small. The formal proof is more involved, and is given in Section A.1.

From Indistinguishability to Entropic Security Proving that indistinguishability implies entropic security is considerably more delicate. Although the statement is a high-entropy version of the equivalence between semantic security and indistinguishability of encryptions due to Goldwasser and Micali [10], the proof techniques are quite different and so we begin with an overview of the main ideas and notation.

The Case of Balanced Predicates We say a function f is *balanced* (w.r.t. X) if it takes on all its possible values with equal probability, i.e. there are $\frac{1}{\mathsf{pred}_f}$ possible values and each occurs with probability pred_f. The reductions we consider are much easier for balanced functions—most of the effort will be in reducing unbalanced functions to balanced ones without losing too much in the prediction probability.

For example, suppose that $g()$ is a balanced *predicate* for distribution X, that is $\Pr[g(X)=0] = \Pr[g(X)=1] = \frac{1}{2}$, and that that $\mathcal{A}$ is an adversary contradicting entropic security for min-entropy

$t = \mathbf{H}_\infty(X)$, that is $\Pr[\mathcal{A}(Y(X)) = g(X)] = \frac{1}{2} + \epsilon$. For $b \in \{0,1\}$, let X_b be the distribution of X conditioned on $g(X) = b$. The adversary's advantage over random guessing in distinguishing $Y(X_0)$ from $Y(X_1)$ is ϵ. However, that same advantage is also a lower bound for the statistical difference. We get:

$$\begin{aligned} \tfrac{1}{2} + \epsilon &= \Pr[\mathcal{A}(Y(X)) = g(X)] \\ &= \Pr[b \leftarrow \{0,1\} : \ \mathcal{A}(Y(X_b)) = b] \leq \tfrac{1}{2} + \tfrac{1}{2}\mathbf{SD}\left(Y(X_0), Y(X_1)\right), \end{aligned}$$

and so the distance between $Y(X_0)$ and $Y(X_1)$ is at least $\epsilon/2$. To see that this contradicts indistinguishability, note that since $g(X)$ is balanced, we obtain X_0 and X_1 by conditioning on events of probability at least $\frac{1}{2}$. Probabilities are at most doubled, and so the min-entropies of both X_0 and X_1 are at most $\mathbf{H}_\infty(X) - 1$.

Balancing Predicates If the predicate $g()$ is not balanced on X, then the previous strategy yields a poor reduction. For example, $\Pr[g(X) = 0]$ may be very small (potentially as small as ϵ). The probabilities in the distribution X_0 would then be a factor of $1/\epsilon$ bigger than their original values, leadding to a loss of min-entropy of $\log(1/\epsilon)$. This argument therefore proves a weak version of Theorem 2.1: (t, ϵ) indistinguishability implies $(t + \log\left(\frac{1}{\epsilon}\right), 2\epsilon)$ entropic security for *predicates*.

This entropy loss is not necessary. We give a better reduction in Section A.1. The idea is that to change the predicate $g()$ into a balanced predicate by flipping the value of the predicate on points on which the original adversary $\mathcal{A}$ performed poorly. By greedily choosing a set of points in $g^{-1}(0)$ of the right size, we show that there exists a balanced predicate $g'()$ on which the same adversary as before has advantage at least $\epsilon/2$, if the adversary had advantage ϵ for the original predicate.

From Predicates to Arbitary Functions In order to complete the proof of Theorem 2.1, we need to show that entropic security for predicates implies entropic security for all functions. The reduction is captured by the following lemma, which states that for every function with a good predictor (i.e. a predictor with advantage at least ϵ), there exists a predicate for which nearly the same predictor does equally well. This is the main technical result of this section.

The reduction uses the predictor $\mathcal{A}(Y(X))$ as a black box, and so we will simply use the random variable $A = \mathcal{A}(Y(X))$.

Lemma 2.2 (Main Lemma). *Let X be any distribution on $\{0,1\}^n$ such that $t \geq \frac{3}{2}\log\left(\frac{1}{\epsilon}\right)$, and let A be any random variable (possibly correlated to X). Suppose there exists a function $f : \{0,1\}^n \to \{0,1\}^*$ such that $\Pr[A = f(X)] \geq \mathsf{pred}_f + \epsilon$. Then there exists a* predicate $g : \{0,1\}^n \to \{0,1\}$ *and an algorithm $B(\cdot)$ such that*

$$\Pr[B(A) = g(X)] \geq \mathsf{pred}_g + \epsilon/4.$$

There are two main steps to proving the lemma:

- If A is a good predictor for an (arbitrary) function $f(\cdot)$, then there is a (almost) *balanced* function $f'(\cdot)$ and a good predictor A's of the form $g(A)$.
- If $f(\cdot)$ is a balanced function (or almost balanced) and A is a good predictor for $f(X)$, then there is a predicate $g(\cdot)$ of the form $g'(f(\cdot))$ such that $g'(A)$ is a good predictor for $g(X)$.

The proof itself is in Section A.2.2.

A More Efficient Reduction Lemma 2.2 says nothing about the running time of $B(\cdot)$—in general, the reduction may yield a large circuit. Nonetheless, we may indeed obtain a polynomial-time reduction for certain functions f. If no value of f occurs with probability more than ϵ^2, then inner product with a random vector provides a good predicate.

Proposition 2.3. *Let X be any random variable distributed in $\{0,1\}^n$. Let $f : \{0,1\}^n \to \{0,1\}^N$ be a function such that $\mathsf{pred}_{f,X} \le \epsilon^2/4$, and let A be a random variable with advantage ϵ at guessing $f(X)$. For $r \in \{0,1\}^N$, let $g_r(x) = r \odot f(x)$. If r is drawn uniformly from $\{0,1\}^N$, then*

$$\mathsf{Exp}_r \left[\Pr[r \odot A = g_r(X)] - \mathsf{pred}_{g_r} \right] \ge \epsilon/4.$$

In particular, there exists r and a $O(N)$-time circuit B such that $\Pr[B(A) = g_r(X)] \ge \mathsf{pred}_{g_r} + \epsilon/4$.

We prove Proposition 2.3 in Section A.2.2, and use it as motivation for the proof of Lemma 2.2.

3 Encryption of High-Entropy Sources

In this section, we discuss the results on entropic security to the encryption of mesages which are guaranteed to come from a high-entropy distribution. Roughly: if the adversary has only a small chance of guessing the message ahead of time, then one can design information-theoretically secure encryption (in the sense of hiding all functions, Definition 1.1) using a much shorter key than is usually possible—making up for the small entropy of the key using the entropy inherent in the message.

3.1 Using Expander Graphs for Encryption

Formally, a symmetric encryption scheme is a pair of randomized maps $(\mathcal{E}, \mathcal{D})$. The encryption takes three inputs, an n-bit message m, a k-bit key κ and r random bits i, and produces a N-bit ciphertext $y = \mathcal{E}(m, \kappa; i)$. Note that the key and the random bits are expected to be uniform random bits, and when it is not necessary to denote the random bits or key explicitly we use either $\mathcal{E}(m, \kappa)$ or $\mathcal{E}(m)$. The decryption takes a key κ and ciphertext $y \in \{0,1\}^N$, and produces the plaintext $m' = \mathcal{D}(y, \kappa)$. The only condition we impose for $(\mathcal{E}, \mathcal{D})$ to be called an encryption scheme is completeness: for all keys κ, $\mathcal{D}(\mathcal{E}(m, \kappa), \kappa) = m$ with probability 1.

In this section, we discuss graph-based encryption schemes and show that graph expansion corresponds to entropically secure encryption schemes.

Graph-based Encryption Schemes Let $G = (V, E)$ be a d-regular graph, and let $N(v, j)$ denote the j-th neighbor of vertex v under some particular labeling of the edges. We'll say the labeling is *invertible* if there exists a map N' such that $N(v, j) = w$ implies $N'(w, j) = v$.

By Hall's theorem, every d-regular graph has an invertible labeling.[7] However, there is a large class of graphs for which this invertibility is much easier to see. The Cayley graph $G = (V, E)$ associated with a group $\mathcal{G}$ and a set of generators $\{g_1, ..., g_d\}$ consists of vertices labeled by elements of $\mathcal{G}$ which are connected when they differ by a generator: $E = \{(u, u \cdot g_i)\}_{u \in V, i \in [d]}$. When the set of generators contains all its inverses, the graph is undirected. For such a graph, the natural labeling is indeed invertible, since $N(v, j) = v \cdot j$ and $N'(w, j) = w \cdot j^{-1}$. All the graphs we discuss in this paper are in fact Cayley graphs, and hence invertibly labeled.

Now suppose the vertex set is $V = \{0,1\}^n$ and the degree is $d = 2^k$, so that the neighbor function N takes inputs in $\{0,1\}^n \times \{0,1\}^k$. Consider the encryption scheme:

$$\mathcal{E}(m, \kappa) = N(m, \kappa). \tag{1}$$

[7] We thank Noga Alon for pointing out this fact. If $G = (V, E)$ is a d-regular undirected graph, consider the bipartite graph with $|V|$ edges on each side and where each edge in E is replaced by the corresponding pair of edges in the bipartite graph. By Hall's theorem, there exist d disjoint matchings in the bipartite graph. These induce an invertible labeling on the original graph.

Notice, $\mathcal{E}$ is a proper encryption scheme if and only if the labeling is invertible. In that case, $\mathcal{D}(y,\kappa) = N'(y,\kappa) = m$. For efficiency, we should be able to compute N and N' in polynomial time. We will show that this encryption scheme is secure when the graph G is a sufficiently good expander. The following definition is standard:

Definition 3.1. *A graph $G = (V, E)$ is a (t, ϵ)-extractor if, for every set S of 2^t vertices, taking a random step in the graph from a random vertex of S leads to a nearly uniform distribution on the whole graph. That is, let U_S be uniform on S, J be uniform on $\{1, ..., d\}$ and U_V be uniform on the entire vertex set V. Then for all sets S of size at least 2^t, we require that:*

$$\mathbf{SD}\left(\ N(U_S, J)\ ,\ U_V\ \right) \leq \epsilon.$$

The usual way to obtain extractors as above is to use good expanders. This is captured by the following lemma.

Lemma 3.1 (Expander smoothing lemma [11]). *A graph G with second largest (normalized) eigenvalue $\lambda \leq \epsilon 2^{-(n-t)/2}$ is a (t, ϵ)-extractor.*

The equivalence between entropic security and indistinguishability (Theorem 2.1) gives us the following result:

Proposition 3.2. *For a 2^k-regular, invertible graph G as above, the encryption scheme $(\mathcal{E}, \mathcal{D})$ given by N, N' is (t, ϵ)-entropically secure if G is a $(t-2, 2\epsilon)$-extractor (in particular, if G has second eigenvalue $\lambda \leq \epsilon \cdot 2^{-(n-t-2)/2}$).*

Proof. By Theorem 2.1, it suffices to show that $(t-2, \epsilon)$-indistinguishability. And this immediately follows from the lemma above and the fact that any min-entropy $(t-2)$ distribution is a mixture of flat distributions. □

We apply this in two ways. First, using optimal expanders (Ramanujan graphs) we obtain the best known construction of entropically-secure encryption schemes (Corollary 3.3). Second, we give a simpler and much stronger analysis of the original scheme of Russell and Wang (Corollary 3.4).

Corollary 3.3. *There exists an efficient deterministic (t, ϵ)-entropically secure scheme with $k = n - t + 2\log\left(\frac{1}{\epsilon}\right) + 2$.*

Proof. We apply Proposition 3.2 to *Ramanujan graphs.* These graphs are optimal for this particular construction: they achieve optimal eigenvalue $\lambda = 2\sqrt{d-1}$ for degree d [16]. The bound on k now follows. □

The main drawback of Ramanujan graphs is that explicit constructions are not known for all sizes of graphs and degrees. However, large families exist (e.g. graphs with $q+1$ vertices and degree $p+1$, where p and q are primes congruent to 1 mod 4). Below we show why the construction from Russell and Wang [22] using small-biased spaces is actually a special case of Proposition 3.2.

Using Small-biased Sets A set S in $\{0,1\}^n$ is *δ-biased* if for all nonzero $\alpha \in \{0,1\}^n$, the binary inner product $\alpha \odot s$ is nearly balanced for s drawn uniformly in S:

$$\Pr_{s \leftarrow S}[\alpha \odot s = 0] \in \left[\frac{1-\delta}{2}, \frac{1+\delta}{2}\right] \quad \text{or, equivalently,} \quad \left|\mathsf{Exp}_{s \leftarrow S}\left[(-1)^{\alpha \odot S}\right]\right| \leq \delta. \tag{2}$$

Alon et al. [1] gave explicit constructions of δ-biased sets in $\{0,1\}^n$ with size $O(n^2/\delta^2)$. Now suppose the δ-biased set is indexed $\left\{s_\kappa | \kappa \in \{0,1\}^k\right\}$. Consider the encryption scheme: $\mathcal{E}(m,\kappa) = m \oplus s_\kappa$. Russell and Wang introduced this scheme and showed that it is (n, ϵ)-entropically secure when $\delta = \epsilon^{3/2}$, yielding a key length of $k = 2\log n + 3\log\left(\frac{1}{\epsilon}\right)$. However, their analysis works only when the message is drawn uniformly from $\{0,1\}^n$.

We propose a different analysis: consider the Cayley graph for $\mathbb{Z}_2^n$ with generators S, where S is δ-biased. This graph has second eigenvalue $\lambda \le \delta$ [19, 2, 5]. Hence, by Proposition 3.2 the scheme above is (t,ϵ)-entropically secure as long as $\delta \le \epsilon 2^{-(n-t-2)/2}$. This gives a version of the Vernam one-time pad for high-entropy message spaces, with key length $k = n-t+2\log n+2\log\left(\frac{1}{\epsilon}\right)+O(1)$. Unlike [22], this works for *all* settings of t, and also improves the parameters in [22] for $n = t$.

Corollary 3.4. *If $\left\{s_\kappa | \kappa \in \{0,1\}^k\right\}$ is a δ-biased set, then the encryption scheme $\mathcal{E}(m,\kappa) = m \oplus s_\kappa$ is (t,ϵ) indistinguishable when $\epsilon = \delta 2^{(n-t-2)/2}$. Using the costruction of [1], this yields a scheme with key length $k = n-t+2\log\left(\frac{1}{\epsilon}\right)+2\log(n)+O(1)$ (for any value of t).*

3.2 A Random Hashing Construction

This section presents a simpler construction of entropically secure encryption based on pairwise independent hashing. Our result generalizes the construction of Russell and Wang [22] for nonuniform sources, and introduces a new variant of the leftover-hash/privacy-amplification lemma [3, 14].

The idea behind the construction is that indistinguishability is the same as extraction from a weak source, except that the extractor must in some sense be invertible: given the key, one must be able to recover the message.

Let $\{h_i\}_{i\in I}$ be some family of functions $h_i : \{0,1\}^k \to \{0,1\}^n$, indexed over the set $I = \{0,1\}^r$. We consider encryption schemes of the form

$$\begin{aligned} \mathcal{E}(m,\kappa;i) &= (\ i,\ m \oplus h_i(\kappa)\ ) \quad \text{(for general functions } h_i\text{), or} && (3)\\ \mathcal{E}'(m,\kappa;i) &= (\ i,\ h_i(m) \oplus \kappa\ ) \quad \text{(when the functions } h_i \text{ are permutations)} && (4)\end{aligned}$$

Notice that this schemes are special low-entropy, probabilistic one-time pads. Decryption is obviously possible, since the description of the function h_i is public. For the scheme to be (t,ϵ)-secure, we will see that it is enough to have $k = n-t+2\log\left(\frac{1}{\epsilon}\right)+2$, and for the function family to be pairwise independent. (This matches the result in Corollary 3.3.) In fact, a slightly weaker condition is sufficient: The following definition was introduced in the context of authentication [15]:

Definition 3.2 (XOR-universal function families). *A collection of functions $\{h_i\}_{i\in I}$ from n bits to n bits is* XOR-universal *if: $\forall a,x,y \in \{0,1\}^n, x \ne y :\quad \Pr_{i\leftarrow I}[h_i(x)\oplus h_i(y) = a] \le \frac{1}{2^n-1}$.*

It is easy to construct XOR-universal families. Any (ordinary) pairwise independent hash family will do, or one can save some randomness by avoiding the "offset" part of constructions of the form $h(x) = ax+b$. Specifically, view $\{0,1\}^n$ as $\mathcal{F} = GF(2^n)$, and embed the key set $\{0,1\}^k$ as a subset of $\mathcal{F}$. For any $i \in \mathcal{F}$, let $h_i(\kappa) = i\kappa$, with multiplication in $\mathcal{F}$. This yields a family of linear maps $\{h_i\}$ with 2^n members. Now fix any $a \in \mathcal{F}$, and any $x,y \in \mathcal{F}$ with $x \ne y$. When i is chosen uniformly from $\{0,1\}^n$, we have $h_i(x)\oplus h_i(y) = i(x-y) = a$ with probability exactly 2^{-n}. If we restrict i to be nonzero, then we get a family of *permutations*, and we get $h_i(x)\oplus h_i(y) = a$ with probability at most $\frac{1}{2^n-1}$.

Proposition 3.5. *If the family $\{h_i\}$ is XOR-universal, then the encryption schemes*

$$\mathcal{E}(m,\kappa;i) = (i, m\oplus h_i(\kappa)) \quad \textit{and} \quad \mathcal{E}'(m,\kappa;i) = (i, h_i(m)\oplus\kappa)$$

are (t,ϵ)-entropically secure, for $t = n-k+2\log\left(\frac{1}{\epsilon}\right)+2$. (However, $\mathcal{E}'$ is a proper encryption scheme only when $\{h_i\}$ is a family of permutations.)

This proposition proves, as a special case, the security of the Russell-Wang construction, with slightly better parameters (their argument gives a key length of $n - t + 3\log\left(\frac{1}{\epsilon}\right)$ since they used 3-wise independent permutations, which are also harder to construct). It also proves the security of the simple construction $\mathcal{E}(m, \kappa; i) = (i, m + i\kappa)$, with operations in $GF(2^n)$.

Proposition 3.5 follows from the following lemma of independent interest, which is closely related to the to the *leftover hash lemma* [12] (also called *privacy amplification*; see, e.g. [3, 4]), and which conveniently handles both the $\mathcal{E}$ and the $\mathcal{E}'$ variants.

Lemma 3.6. *If A, B are independent random variables such that $\mathbf{H}_\infty(A)+\mathbf{H}_\infty(B) \geq n+2\log\left(\frac{1}{\epsilon}\right)+1$, and $\{h_i\}$ is a XOR-universal family, then $\mathbf{SD}\left(\langle i,\ h_i(A) \oplus B\rangle\ ,\ \langle\ i,\ U_n\rangle\right) \leq \epsilon$, where U_n and i are uniform on $\{0,1\}^n$ and $\mathcal{I}$.*

Proof of Lemma 3.6. Consider the collision probability of $(i,\ h_i(A) \oplus B)$. A collision only occurs if the same function h_i is chosen both times. Conditioned on that, one obtains a collision only if $h_i(A) \oplus h_i(A') = B \oplus B'$, for A', B' i.i.d. copies of A, B. We can use the XOR-universality to bound this last term:

$$\begin{aligned}\Pr[(i,\ h_i(A) \oplus B) = (i,\ h_i(A') \oplus B')] \\ = \Pr[i = i']\Big(\Pr[B = B'] \cdot \Pr[h_i(A) = h_i(A')] \\ + \sum_{a \neq 0} \Pr[B \oplus B' = a] \cdot \Pr[h_i(A) \oplus h_i(A') = a]\Big)\end{aligned} \tag{5}$$

Now let $t_a = \mathbf{H}_2(A)$, $t_b = \mathbf{H}_2(B)$. For $a \neq 0$, we have $\Pr[h_i(A) \oplus h_i(A') = a] \leq 1/(2^n - 1)$, by the conditions on $\{h_i\}$. On the other hand, by a union bound we have

$$\Pr[h_i(A) = h_i(A')] \leq \Pr[A = A'] + \frac{1}{2^n - 1} \leq 2^{-t_a} + \frac{1}{2^n - 1}$$

Hence, Eqn. 5 reduces to

$$\begin{aligned}\frac{1}{|\mathcal{I}|}\left(2^{-t_b}\left(2^{-t_a} + \frac{1}{2^n - 1}\right) + \frac{1}{2^n - 1}\left(\sum_{a \neq 0} \Pr[B \oplus B' = a]\right)\right) \\ \leq \frac{1}{|\mathcal{I}|2^n}\left(1 + 2^{n - t_a - t_b} + 2^{-t_b} + \frac{2}{2^n - 1}\right)\end{aligned}$$

Now $2^{n-t_a-t_b} \leq \epsilon^2/2$ by assumption, and we also have $2^{-n} \leq 2^{-t_b} \leq \epsilon^2/2$, since $t_a, t_b \leq n$ and $t_a + t_b \geq n + 2\log\left(\frac{1}{\epsilon}\right)$ (similarly, $n \geq 2\log\left(\frac{1}{\epsilon}\right)$). Hence Eqn. 5 reduces to $(1 + 2\epsilon^2)/|\mathcal{I}|2^n$. As we mentioned, any distribution on a finite set S with collision probability $(1 + 2\epsilon^2)/|S|$ is at statistical distance at most ϵ from the uniform distribution [14]. Thus, $(i, h_i(A) \oplus B)$ is ϵ-far from uniform. □

Note that the lemma gives a special "extractor by XOR" which works for product distributions $A \times B$ with at least n bits on min-entropy between them.

3.3 Lower Bounds on the Key Length

Proposition 3.7. *Any encryption scheme which is (t, ϵ)-entropically secure for inputs of length n requires a key of length at least $n - t$.*

Proof. We can reduce our entropic scheme to Shannon-secure encryption of strings of length $n - t + 1$. Specifically, for every $w \in \{0,1\}^{n-t+1}$, let M_w be the uniform over strings with w as a prefix, that is the set $\{w\} \times \{0,1\}^{t-1}$. Since M_w has min-entropy $t-1$, any pair of distributions $\mathcal{E}(M_w), \mathcal{E}(M_{w'})$ are indistinguishable, and so we can use $\mathcal{E}()$ to encrypt strings of length $n-t+1$. When $\epsilon < 1/2$, we must have key length at least $(n-t+1)-1 = n-t$ by the usual Shannon-style bound (the loss of 1 comes from a relaxation of Shannon's bounds to statistical security). □

Bounds for Public-Coin Schemes via Extractors In the constructions of Russell and Wang and that of Section 3.1 and Section 3.2, the randomness used by the encryption scheme (apart from the key) is sent *in the clear* as part of the ciphertext. That is, $\mathcal{E}(m,\kappa;i) = (i, \mathcal{E}'(m.\kappa;i))$. For these types of schemes, called *public-coin* schemes, the intuitive connection between entropic security and extraction from weak sources is pretty clear: encryption implies extraction. As a result, lower bounds on extractors [21] apply, and show that our construction is close to optimal.

Proposition 3.8. *Any* public-coin*, (t,ϵ)-entropically secure encryption has key length $k \geq n-t+\log\left(\frac{1}{\epsilon}\right) - O(1)$.*

To prove the result, we first reduce to the existence of extractors:

Lemma 3.9. *Let $(\mathcal{E},\mathcal{D})$ be a* public-coin*, (t,ϵ)-entropically secure encryption scheme with message length n, key length k and r bits of extra randomness. Then there exists an extractor with seed length $k+r$, input length n and output length $n+r-\log\left(\frac{1}{\epsilon}\right)$, such that for any input distribution of min-entropy $t+1$, the output is within distance 3ϵ of the uniform distribution.*

Proof. We combine three observations. First, when U is uniform over all messages in $\{0,1\}^n$, the entropy of the distribution $\mathcal{E}(U)$ must be high. Specifically: $\mathbf{H}_\infty(\mathcal{E}(U)) = n+r$. To see this, notice that there is a function $(\mathcal{D})$ which can produce R,K,U from $K, \mathcal{E}(U,K;R)$. Since the triple (R,K,U) is uniform on $\{0,1\}^{r+k+n}$, it must be that $(K, \mathcal{E}(U,K))$ also has min-entropy $r+k+n$, i.e. that any pair (κ, c) appears with probability at most $2^{-(n-k-r)}$. Summing over all 2^k values of κ, we see that any ciphertext value c appears with probability at most $\sum_\kappa 2^{-n-r-k} = 2^{-n-r}$, as desired.

The second observation is that there is a deterministic function ϕ which maps ciphertexts into $\{0,1\}^{n+r-\log\left(\frac{1}{\epsilon}\right)}$ such that $\phi(\mathcal{E}(U))$ is within distance ϵ of the uniform distribution. In general, any *fixed* distribution of min-entropy t can be mapped into $\{0,1\}^{t-\log(1/\epsilon)}$ so that the result is almost uniform (Simply assign elements of the original distribution one by one to strings in $\{0,1\}^{t-\log(1/\epsilon)}$, so that at no time do two strings have difference of probability more than 2^{-t}. The total variation from uniform will be at most $2^{t-\log(1/\epsilon)} \cdot 2^{-t} = \epsilon$.). Note that ϕ need not be efficiently computable, even if both $\mathcal{E}$ and $\mathcal{D}$ are straightforward. This doesn't matter, since we are after a combinatorial contradiction.

Finally, by Theorem 2.1, for all distributions of min-entropy $t-1$, we have $\mathbf{SD}\left(\mathcal{E}(U), \mathcal{E}(M)\right) \leq 2\epsilon$, and so $\mathbf{SD}\left(\phi(\mathcal{E}(U)), \phi((\mathcal{E}(M))\right) \leq 2\epsilon$. By the triangle inequality, $\phi(\mathcal{E}(M))$ is within 3ϵ of the uniform distribution on $n+r-\log\left(\frac{1}{\epsilon}\right)$ bits, proving the lemma. □

We can now apply the lower bound of Radhakrishnan and Ta-Shma [21], who showed that any extractor for distributions of min-entropy t, distance parameter δ and d extra random bits, can extract at most $t+d-2\log(1/\delta)+O(1)$ nearly random bits. From Lemma 3.9, we get and extractor for min-entropy $t+1$, $\delta = 3\epsilon$, $k+r$ extra random bits, and output length $n+r-\log(1/\epsilon)$. Thus, $n+r-\log(1/\epsilon)$ is at most $t+1+k+r-2\log(1/\epsilon)+O(1)$, which immediately gives us Proposition 3.8.

Remark 3.1. We do not lose $\log(1/\epsilon)$ in the output length in Lemma 3.9 when the encryption scheme in indistinguishable from the uniform distribution (i.e., ciphertexts look truly random). For such public-coin schemes, we get $k \geq n-t+2\log\left(\frac{1}{\epsilon}\right) - O(1)$. Since all of our constructions are of this form, their parameters cannot be improved at all. In fact, we conjecture that $k \geq n-t+2\log\left(\frac{1}{\epsilon}\right) - O(1)$ for *all* entropically-secure schemes, public-coin or not.

Acknowledgements

We are grateful to many friends for helpful discussions on this work. We especially thank Noga Alon, Leonid Reyzin, Madhu Sudan, Salil Vadhan and Avi Wigderson for their insights.

References

[1] Noga Alon, Oded Goldreich, Johan Håstad, René Peralta: Simple Constructions of Almost k-Wise Independent Random Variables. FOCS 1990: 544-553

[2] Noga Alon and Yuval Roichman. Random Cayley graphs and expanders. *Random Structures & Algorithms* 5 (1994), 271–284.

[3] C. Bennett, G. Brassard, and J. Robert. Privacy Amplification by Public Discussion. *SIAM J. on Computing*, **17**(2), pp. 210–229, 1988.

[4] C. Bennett, G. Brassard, C. Crépeau, and U. Maurer. Generalized Privacy Amplification. *IEEE Transactions on Information Theory*, **41**(6), pp. 1915-1923, 1995.

[5] Eli Ben-Sasson, Madhu Sudan, Salil P. Vadhan, Avi Wigderson: Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. STOC 2003: 612-621

[6] R. Canetti. Towards realizing random oracles: Hash functions that hide all partial information. In *Crypto 1997*.

[7] R. Canetti, D. Micciancio, O. Reingold. Perfectly One-Way Probabilistic Hash Functions. In *Proc. 30th ACM Symp. on Theory of Computing*, 1998, pp. 131–140.

[8] Ran Canetti. Universally Composable Security: A New Paradigm for Cryptographic Protocols. *Proc. IEEE Symp. on Foundations of Computer Science*, 2001, pp. 136-145.

[9] T. Cover, J. Thomas. *Elements of Information Theory.* Wiley series in telecommunication, 1991, 542 pp.

[10] S. Goldwasser and S. Micali. Probabilistic encryption. *JCSS*, **28**(2), pp. 270–299, April 1984.

[11] Oded Goldreich, Avi Wigderson: Tiny families of functions with random properties: A quality-size trade-off for hashing. Random Structures and Algorithms 11(4): 315-343 (1997)

[12] J. Håstad, R. Impagliazzo, L. Levin, M. Luby. A Pseudorandom generator from any one-way function. In *Proc. 21st ACM Symp. on Theory of Computing*, 1989.

[13] Jonathan Herzog. *Computational Soundness for Standard Assumptions of Formal Cryptography.* Ph.D. Thesis, Massachusetts Institute of Technology, May 2004.

[14] R. Impagliazzo and D. Zuckerman. How to Recycle Random Bits. In *Proc. 30th IEEE Symp. on Foundations of Computer Science*, 1989.

[15] H. Krawczyk. LFSR-Based Hashing and Authentication. In *Proc. CRYPTO '94*, p. 129–139, 1994.

[16] A. Lubotzky, R. Phillips, P. Sarnak: Ramanujan graphs. Combinatorica 8(3): 261-277 (1988).

[17] U. Maurer. Conditionally-Perfect Secrecy and a Provably-Secure Randomized Cipher. *J. Cryptology*, **5**(1), pp. 53–66, 1992.

[18] U. Maurer. Secret Key Agreement by Public Discussion. *IEEE Trans. on Info. Theory*, 39(3):733–742, 1993.

[19] J. Naor, M. Naor.Small-Bias Probability Spaces: Efficient Constructions and Applications.In *SIAM J. Comput.* 22(4): 838-856 (1993).

[20] N. Nisan, D. Zuckerman. Randomness is Linear in Space. In *JCSS*, **52**(1), pp. 43–52, 1996.

[21] J. Radhakrishnan and A. Ta-Shma. Tight bounds for depth-two superconcentrators. In *Proc. 38th IEEE Symp. on Foundations of Computer Science*, 1997, pp. 585–594.

[22] A. Russell and Wang. How to Fool an Unbounded Adversary with a Short Key. In *Advances in Cryptology — EUROCRYPT 2002.*

[23] C. Shannon. Communication Theory of Secrecy systems. In *Bell Systems Technical J.*, 28:656–715, 1949. Note: The material in this paper appeared originally in a confidential report 'A Mathematical Theory of Cryptography', dated Sept. 1, 1945, which has now been declassified.

A Proofs of Equivalences

A.1 From Entropic Security to Indistinguishability

Lemma A.1. *(t,ϵ)-entropic security for predicates implies $(t-1,4\epsilon)$-indistinguishability.*

Proof. It is sufficient to prove indistinguishability for all distributions which are uniform on some set of 2^{t-1} points. To see why, recall that any distribution of min-entropy at least $t-1$ can be written as a convex combination of such *flat* distributions. If $X_0 = \sum \lambda_{0,i} X_{0,i}$ and $X_1 = \sum_j \lambda_{1,j} X_{1,j}$, where the $X_{0,i}$ and $X_{1,j}$ are all flat distributions, then the statistical distance $\mathbf{SD}\left(Y(X_0), Y(X_1)\right)$ is bounded above by $\sum_{i,j} \lambda_{0,i}\lambda_{1,j}\mathbf{SD}\left(Y(X_{0,i}), Y(X_{1,j})\right)$ (by the triangle inequality). If each of the pairs $Y(X_{0,i}), Y(X_{1,j})$ has distance at most ϵ, then the entire sum will be bounded by ϵ.

Now let X_0, X_1 be any two flat distributions over *disjoint* sets of 2^{t-1} points each (we will deal with non-disjoint sets below), and let X be an equal mixture of the two. That is, to sample from X, flip a fair coin B, and sample from X_B. Take g to be any predicate which is 0 for any sample from X_0 and 1 for any sample from X_1. A good predictor for g will be the adversary $\mathcal{A}$ who, given a string y as input, guesses as follows:

$$\mathcal{A}(y) = \begin{cases} 0 & \text{if } y \text{ is more likely under the distribution } Y(X_0) \text{ than under } Y(X_1) \\ 1 & \text{otherwise} \end{cases}$$

By the definition of statistical difference, this adversary guesses the predicate with probability exactly:

$$\Pr\left[\mathcal{A}(Y(X)) = B = g(X)\right] = \tfrac{1}{2} + \tfrac{1}{2}\mathbf{SD}\left(Y(X_0), Y(X_1)\right). \tag{6}$$

We can now apply the assumption that $Y()$ is (t,ϵ)-entropically secure to bound $\mathbf{SD}\left(Y(X_0), Y(X_1)\right)$. First, for any random variable G over $\{0,1\}$ which is independent of X, the probability that $G = g(X)$ is exactly $\frac{1}{2}$. Now the distribution X has min-entropy t by construction, and so by entropic security the probability that $\mathcal{A}(y)$ can guess $g(X)$ is bounded:

$$\Pr[\mathcal{A}(Y(X)) = g(X)] \leq \max_G \left\{\Pr[G = g(X)]\right\} + \epsilon = \tfrac{1}{2} + \epsilon. \tag{7}$$

Combining the last two equations, the statistical difference $\mathbf{SD}\left(Y(X_0), Y(X_1)\right)$ is at most 2ϵ. This takes care of the case where X_0 and X_1 have disjoint supports.

To get the general indistinguishability condition, fix any $\tilde{X}_0$ as above (flat on 2^{t-1} points). For any other flat distribution $\tilde{X}_1$, there is some third flat distribution X' which is disjoint from both

$\tilde{X}_0$ and $\tilde{X}_1$. By the previous reasoning, both $\mathbf{SD}\left(Y(\tilde{X}_0), Y(X')\right)$ and $\mathbf{SD}\left(Y(X'), Y(\tilde{X}_1)\right)$ are less than 2ϵ. By the triangle inequality $\mathbf{SD}\left(Y(X_0), Y(X_1)\right) \leq 4\epsilon$ (a more careful proof avoids the triangle inequality and gives distance 2ϵ even when the supports of X_0, X_1 overlap. □

A.2 From Indistinguishability to Entropic Security

A.2.1 Entropic Security for Predicates

Lemma A.2. *$(t-2, 2\epsilon)$-indistinguishability implies (t, ϵ)-entropic security for* ***predicates*** *for* $t \geq 2$.

Proof. Suppose that the scheme is not (t, ϵ)-entropically secure. That is, there is a message distribution X with min-entropy at least t, a predicate g and an adversary $\mathcal{A}$ such that

$$\Pr[\mathcal{A}(Y(X)) = g(X)] > \epsilon + \max_{i=0,1}\{\Pr[g(X) = i]\} \tag{8}$$

We wish to choose two distributions of min-entropy $t-2$ and use the adversary to distinguish them, thus contradicting indistinguishability. It's tempting to choose the sets $g^{-1}(0)$ and $g^{-1}(1)$, since we know the adversary can predict g reasonably well. That attempt fails because one of the pre-images $g^{-1}(0), g^{-1}(1)$ might be quite small, leading to distributions of low min-entropy. Instead, we partition the support of X into sets of (almost) equal measure, making sure that the smaller of $g^{-1}(0)$ and $g^{-1}(1)$ is entirely contained in one partition.

Now let:

$$\begin{aligned} p &= \Pr[h(X) = 1] \\ q_0 &= \Pr[\mathcal{A}(Y(X)) = 1 | g(X) = 0] \\ q_1 &= \Pr[\mathcal{A}(Y(X)) = 1 | g(X) = 1] \end{aligned}$$

Suppose without loss of generality that $p \geq 1/2$, i.e. that $g(X) = 1$ is more likely than, or as likely as, $g(X) = 0$ (if $p < 1/2$, we can just reverse the roles of 0 and 1). The violation of entropic security (Eq. 8) can be re-written:

$$pq_1 + (1-p)(1-q_0) > p + \epsilon$$

In particular, $p - pq_1 > 0$ so we get:

$$(1-p)(q_1 - q_0) > \epsilon \tag{9}$$

Now we wish to choose two distributions A, B, each of min-entropy $t-2$. For now, fix any set $\mathcal{S} \subseteq g^{-1}(1)$, where $g^{-1}(1) = \{m \in \{0,1\}^n | g(m) = 1\}$. We make the choice of § more specific below. Let $A_{\mathcal{S}}$ be the conditional distribution of X conditioned on $X \in \mathcal{S}$, and let $B_{\mathcal{S}}$ be distributed as X conditioned on $X \in \{0,1\}^n \setminus \mathcal{S}$. That is, $A_{\mathcal{S}}$ and $B_{\mathcal{S}}$ have disjoint supports and the support of $B_{\mathcal{S}}$ covers $g^{-1}(0)$ entirely.

The first property we will need from $\mathcal{S}$ is that it split the mass of X somewhat evenly. If the probability mass p' of $\mathcal{S}$ under X was exactly $1/2$, then the min-entropies of $A_{\mathcal{S}}$ and $B_{\mathcal{S}}$ would both be exactly $t-1$. Depending on the distribution X, it may not be possible to have such an even split. Nonetheless, we can certainly get $\frac{1}{2} \leq p' < \frac{1}{2} + 2^{-t}$, simply by adding points one at a time to § until it gets just below 1/2. The order in which we add the points is not important. For $t > 2$ (which is a hypothesis of this proof), we get $\frac{1}{2} \geq p' \geq \frac{3}{4}$. Hence, we can choose $\mathcal{S}$ so that the min-entropies of $A_{\mathcal{S}}$ and $B_{\mathcal{S}}$ are both at least $t-2$.

We will also need that $\mathcal{S}$ have other properties. For every point x in the support of X, we define $q_x = \Pr[\mathcal{A}(Y(x)) = 1]$. The average over $x \leftarrow X$, restricted to $g^{-1}(1)$, of q_x is exactly q_1, that is

$$\mathsf{Exp}_{x \leftarrow X}[q_x] = q_1$$

If we now the choose the set $\mathcal{S}$ greedily, always adding points which maximize q_x, we are guaranteed that the average over X, conditioned on $X \in \mathcal{S}$, is at least q_1. That is, there exists a choice of S with mass $p' \in [\frac{1}{2}, \frac{3}{4}]$ such that

$$\Pr[\mathcal{A}(Y(A_{\mathcal{S}})) = 1] = \mathsf{Exp}_{x \leftarrow A_{\mathcal{S}}}[q_x] \geq q_1.$$

We can also now compute the probability that $\mathcal{A}(Y(B_{\mathcal{S}}))$ is 1:

$$\Pr[\mathcal{A}(Y(B_{\mathcal{S}})) = 1] = \frac{1-p}{1-p'}q_0 + \frac{p-p'}{1-p'}\Pr[\mathcal{A}(Y(X)) = 1 | X \notin \mathcal{S} \text{ and } g(X) = 0]$$

Now $\Pr[\mathcal{A}(Y(X)) = 1 | X \notin \mathcal{S}$ and $g(X) = 0]$ is at most q_1 (since by the greedy construction of $\mathcal{S}$, this is the average over elements in $g^{-1}(1)$ with the lowest values of q_m). Using $\mathcal{A}$ as a distinguisher for the distributions $Y(A_{\mathcal{S}})$ and $Y(B_{\mathcal{S}})$, we get:

$$\left|\ \Pr\left[\mathcal{A}(Y(A_{\mathcal{S}})) = 1\right] - \Pr\left[\mathcal{A}(Y(B_{\mathcal{S}})) = 1\right]\ \right| \geq q_1 - \frac{1-p}{1-p'}q_0 - \frac{p-p'}{1-p'}q_1 = \frac{1-p}{1-p'}(q_1 - q_0)$$

Since entropic security is violated (Eq. 9), we have $(1-p)(q_1-q_0)/(1-p') > \epsilon/(1-p')$. By construction, we have $p' > \frac{1}{2}$ so the advantage of the predictor is at least 2ϵ, that is:

$$\mathbf{SD}\left(Y(A_{\mathcal{S}}), Y(B_{\mathcal{S}})\right) \geq \left|\ \Pr\left[\mathcal{A}(Y(A_{\mathcal{S}})) = 1\right] - \Pr\left[\mathcal{A}(Y(B_{\mathcal{S}})) = 1\right]\ \right| \geq 2\epsilon$$

Since A and B each have min-entropy at least $t-2$, this contradicts $(t-2, 2\epsilon)$-indistinguishability, completing the proof. □

A.2.2 From Predicates to General Functions

This section contains the proofs of Lemma 2.2 and Proposition 2.3. We begin with Proposition 2.3, since the proof is straightforward and provides some intuition for the proof of Lemma 2.2.

Proof of Proposition 2.3. We can calculate the expected advantage almost directly. Note that conditioned on the event $A = f(X)$, the predictor $r \odot A$ always agrees with $g_r(X)$. When $A \neq f(X)$, they agree with probability exactly $\frac{1}{2}$. Hence, we have

$$\mathsf{Exp}_r\left[\Pr[r \odot A = g_r(X)]\right] = \frac{1}{2} + \frac{1}{2}\Pr[A = f(X)] \geq \frac{1}{2}(1 + \mathsf{pred}_f + \epsilon)$$

We must still bound the expected value of pred_{g_r}. Let $r_z = (-1)^{z \odot r}$. For any particular, r, we can compute pred_{g_r} as $\frac{1}{2} + \frac{1}{2}\left|\sum_z p_z r_z\right|$. Using the fact $\mathsf{Exp}\left[|Z|\right] \leq \sqrt{\mathsf{Exp}\left[Z^2\right]}$ for any random variable Z, we get:

$$\mathsf{Exp}_r\left[\mathsf{pred}_{g_r}\right] = \frac{1}{2} + \frac{1}{2}\mathsf{Exp}_r\left[\left|\sum_z p_z r_z\right|\right] \leq \frac{1}{2} + \frac{1}{2}\sqrt{\mathsf{Exp}_r\left[\left(\sum_z p_z r_z\right)^2\right]}$$

By pairwise independence of the variables r_z, we have $\mathsf{Exp}\left[r_z r_a\right]$ is 1 if $z = a$ and 0 otherwise.

$$\mathsf{Exp}_r\left[\mathsf{pred}_{g_r}\right] \leq \frac{1}{2} + \frac{1}{2}\sqrt{\sum_z p_z^2} \leq \frac{1}{2} + \frac{1}{2}\sqrt{\mathsf{pred}_f}.$$

The last inequality holds since pred_f is the maximum of the values p_z, and the expression $\sum_z p_z^2$ is maximized when $p_z = \mathsf{pred}_f$ for all z (note that this sum is the collision probability of $f(X)$). Combining the two calculations we have

$$\mathsf{Exp}_r\left[\Pr[r \odot A = g_r(X)] - \mathsf{pred}_{g_r}\right] \geq \frac{1}{2}\left(\mathsf{pred}_f + \epsilon - \sqrt{\mathsf{pred}_f}\right)$$

Using the hypothesis that $\mathsf{pred}_f \leq \epsilon^2/4$, we see that the expected advantage is at least $\epsilon/4$. □

We now turn to the proof of Lemma 2.2. It is tempting, as before, to consider predicates of the form $g(x) = g'(f(x))$ (this is certainly the form of the predicates given by Proposition 2.3). This approach cannot work in general: suppose that $Z = f(X)$ takes on values in $\{0, 1, 2\}$ with equal probability, and suppose that A takes the value of $f(X)$ with probability $1/3 + \epsilon$, and each of the other two values with probability $1/3 - \epsilon/2$. Now any predicate of Z takes on some value with probabilty at least $2/3$. A straightforward calculation shows that no matter what value of A is observed, the best strategy is to guess the more likely value of the predicate. Hence, to prove Lemma 2.2 we'll have to consider a richer set of predicates.

Nonetheless, in the special case where f is *balanced* over an even number of outputs, we can consider the simpler predicates of the previous proof. We say $f : \{0,1\}^n \to \{1, ..., F\}$ is δ-far from balanced with respect to X if for every value $z \in [F] = \{1, ..., F\}$ we have $|p_z - 1/F| \le \delta$. Sub-Lemma A.3 shows that essentially the same approach as before works for a balanced function; that is, it is sufficient to choose a random *balanced* predicate.

SubLemma A.3 (Nearly balanced functions). *Suppose F is even and $f : \{0,1\}^n \to [F]$ is δ-almost-balanced. If $\Pr[A = f(X)] \ge \mathsf{pred}_f + \epsilon$, then there is a predicate $g(x) = g'(f(x))$ such that*

$$\Pr[g'(A) = g(X)] \ge \mathsf{pred}_g + \epsilon/2 - \delta\sqrt{F}.$$

In particular, when $F \le 2/\epsilon$ and $\delta \le \epsilon^{3/2}/8$ the predictor $g'(A)$ has advantage at least $\epsilon/4$ over random guessing.

Proof. It is sufficient to consider a *random* predicate $G' : [F] \to \{-1, +1\}$ subject to the constraint that $G'(z) = -1$ on exactly half the elements of $[F]$. (The constraint can be satisfied F is even.) As in the proof of Proposition 2.3, we will compute the expected prediction probability of $G'(A)$ and the expectation of $\mathsf{pred}_{G'}$ separately.

We first compute the expected probability that $G'(A) = G'(f(X))$. Conditioned on the event $A = f(X)$, we always have $G'(A) = G'(f(X))$, and conditioned on $A \neq f(X)$, we have $G'(A) = G'(f(X))$ with probability $\frac{1}{2} - \frac{1}{2(F-1)}$ (the difference from $\frac{1}{2}$ comes from the fact that we choose G' only from functions which are balanced on $[F]$).

Let $\hat{p} = \Pr[A = f(X)]$. The expected prediction probability is given by

$$\mathsf{Exp}_{G'}\left[\Pr[G'(A) = G'(f(X))]\right] = \hat{p} + (1-\hat{p})(\frac{1}{2} - \frac{1}{2(F-1)}) = \frac{1}{2} + \frac{1}{2}(\hat{p} - \frac{1-\hat{p}}{F-1}).$$

By hypothesis $\hat{p} \ge \mathsf{pred}_f + \epsilon \ge 1/F + \epsilon$. Simplifying, we get $\mathsf{Exp}_{G'}\left[\Pr[G'(A) = G'(f(X))]\right] \ge \frac{1}{2} + \epsilon/2$.

We can also compute the expectation of pred_G (as in the proof of Proposition 2.3). Note that if f is perfectly balanced, $\mathsf{pred}_{G'}$ is always exactly $1/2$. More generally, for each z, let $\delta_z = p_z - \frac{1}{2}$ (recall that $|\delta_z| \le \delta$ by hypothesis). Since G' is always balanced on $[F]$, for any particular g' we have $\mathsf{pred}_{g'} = \frac{1}{2} + \frac{1}{2}\left|\sum_z \delta_z g'(z)\right|$ (using the convention that g' maps into $\{\pm 1\}$). In expectation, we can apply the inequality $\mathsf{Exp}\left[|Z|\right] \le \sqrt{\mathsf{Exp}\left[Z^2\right]}$ to get:

$$\mathsf{Exp}_{G'}\left[\mathsf{pred}_{G'}\right] \le \frac{1}{2} + \frac{1}{2}\sqrt{\mathsf{Exp}_{G'}\left[\left(\sum_z \delta_z G'(z)\right)^2\right]} = \frac{1}{2} + \frac{1}{2}\sqrt{\sum_{z,z'} \delta_z \delta_{z'} \mathsf{Exp}_{G'}\left[G'(z)G'(z')\right]}.$$

We know that $\mathsf{Exp}_{G'}\left[G'(z)G'(z')\right]$ is 1 for $z = z'$ and is $\frac{-1}{F-1}$ otherwise. Using $|\delta_z| \le \delta$ we get:

$$\mathsf{Exp}_{G'}\left[\mathsf{pred}_{G'}\right] \le \frac{1}{2} + \frac{1}{2}\sqrt{\sum_z \delta^2 + \sum_{z \neq z'} \frac{\delta^2}{F-1}} \le \frac{1}{2} + \frac{1}{2}\delta\sqrt{2F}.$$

The expectation of $\Pr[G'(A) = G'(f(X))] - \mathsf{pred}_{G'}$ is at least $\epsilon/2 - \delta\sqrt{F}$, as desired. □

SubLemma A.4 (Balancing Functions). *Let f be any function such that* $\Pr[A = f(X)] \geq \mathsf{pred}_f + \epsilon$*. If* $\mathbf{H}_\infty(X) \geq \log(1/\delta)$*, then there is a function f' such that*

1. *f' takes values in $[F]$, for $F \leq \min\left\{\frac{2}{\mathsf{pred}_f}, \frac{4}{\epsilon}\right\} + 2$, and*
2. *f' is δ-almost-balanced.*
3. *$\exists B(\cdot)$ such that* $\Pr[B(A) = f'(X)] \geq \mathsf{pred}_{f'} + \epsilon/4$.

We can prove Sub-Lemma A.4 using two claims: the first reduces the number of possible outputs simply by "bucketing" certain outputs together. The second claim shows that a function with not too many output can be made almost perfectly balanced, as long as the entropy of X is high enough.

Claim A.5. *Let f be any function such that* $\Pr[A = f(X)] \geq \mathsf{pred}_f + \epsilon$*. Then there's a function b such that $f'(x) = b(f(x))$ satisfies $\mathsf{pred}_{f'} \leq \mathsf{pred}_f + \epsilon/2$, and such that f' takes values in $[F]$, for $F \leq \min\left\{\frac{2}{\mathsf{pred}_f}, \frac{4}{\epsilon}\right\} + 2$.*

Proof. We can gradually reduce the number of possible values f can take without decreasing the advantage of the predictor. Let $p_z = \Pr[f(X) = z]$. If there are two values z, z' such that both p_z and p'_z are at most $\mathsf{pred}_f/2 + \epsilon/4$, then we can identify those two outputs. Note that the combined value has probability at most $\mathsf{pred}_f + \epsilon/2$. We can continue to combine pairs of values with combined mass at most $\mathsf{pred}_f + \epsilon/2$ until there is at most one value z with mass less than $\mathsf{pred}_f/2 + \epsilon/4$.

Let $F = \left\lceil \min\left\{\frac{2}{\mathsf{pred}_f}, \frac{4}{\epsilon}\right\} \right\rceil + 1$. At the end of these successive combinations, we will have at most F different outputs remaining. We can thus summarize the previous steps in a single function b with range $[F]$ such that $b(z) = b(z')$ if and only if the outputs z and z' we're identified at some stage. This b satifies the conditions of the theorem: by contruction, we have $\Pr[b(f(X)) = w]$ is at most $\mathsf{pred}_f + \epsilon/4$ for any value w. Moreover, $\Pr[b(A) = b(f(X))] \geq \Pr[A = f(X)] \geq \mathsf{pred}_f + \epsilon$. □

Claim A.6. *Let $f : \{0,1\}^n \to [F]$ be any function such that* $\Pr[A = f(X)] \geq \mathsf{pred}_f + \epsilon$*. If* $\mathbf{H}_\infty(X) \geq \log(1/\delta)$*, then there is a function $f' : \{0,1\}^n \to [F]$ such that*

1. *f' is δ-almost-balanced, and*
2. $\Pr[A = f'(X)] \geq \frac{1}{F} + \frac{1/F}{\mathsf{pred}_f} \cdot \epsilon$.

In particular, if $F \leq \frac{2}{\mathsf{pred}_f}$, then the advantage of A at predicting $f'(X)$ is $\epsilon/2$.

Proof. We may imagine the function f as dividing the points $x \in \{0,1\}^n$ into F buckets. Our goal is to move some of these points between buckets so that all the buckets have approximately the same size. Since no point has mass more than δ, we will be able to have all buckets with mass within δ of $1/F$.

The problem is that moving points between buckets will (in general) decrease the chances that A will predict the function (i.e. bucket identifier) accurately. Fortunately, we're interested in the difference between the predicition probability and the maximum bucket size. As long as the two decrease proportionately, then A will remain a good predictor for the modified function.

We now formalize our intuition. Let $p_z = \Pr[f(X) = z]$, and let $S \subseteq [F]$ be the set of z such that $p_z \geq 1/F$. We will change the function f on inputs x such that $f(x) \in S$, and assign instead a value not in S. We keep moving points as long as some bucket remains with mass above $1/F + \delta$. Note that buckets in S will all have mass in $[1/F, 1/F + \delta]$ at the end of this process.

Consider a large bucket given by value z. To decide which points to move out of the bucket, let $w_x = \Pr[A = f(x)|X = x]$, and let $q_z = \Pr[A = f(X)|f(X) = z]$. The value q_z is the average of w_x over all x in the bucket given by z:

$$q_z = \mathsf{Exp}_{X|f(X)=z}\left[w_X\right]$$

By always moving points with the smallest possible values of w_x, we can ensure that the average w_x in the bucket always remains at least q_z. Specifically, let f' be the new function obtained by balancing the buckets, and let $q'_z = \Pr[A = z | f'(X) = z]$. Then by moving x's with small w_x we can ensure that $q'_z \geq q_z$ for all $z \in S$.

At the end of this process we will have

$$\Pr[A = f'(X)] \geq \sum_{z \in S} \frac{1}{F} q'_z + \sum_{z \notin S} p_z q_z \geq \frac{1}{F} \sum_{z \in S} q_z + \sum_{z \notin S} p_z q_z$$

(The contribution of a bucket not in S can be bounded below by $p_z q_z$, since it's contribution to the prediction probability can only increase with re-balancing.)

The original prediction probability was $\sum_z p_z q_z$. Thus the coefficients of the q_z in the new success probability have gone down by a factor of at most $\frac{1/F}{\mathsf{pred}_f}$ (that is, only coefficients in S have changed, and those have decreased from at most pred_f to at least $1/F$). Hence, we have

$$\frac{\Pr[A = f'(X)]}{\Pr[A = f(X)]} \geq \frac{1/F}{\mathsf{pred}_f}$$

Thus the new probability is at least $\frac{1}{F} + \frac{1/F}{\mathsf{pred}_f} \epsilon$, as desired. □

Combining the lemmas above proves Lemma 2.2.